PLOS ONE

Structural Modeling and *In Silico* Analysis of Human Superoxide Dismutase 2

Mariana Dias Castela de Carvalho, Joelma Freire De Mesquita*

Bioinformatics and Computational Biology Group, Department of Genetics and Molecular Biology, Federal University of Rio de Janeiro State, Rio de Janeiro, Brazil

Abstract

Aging in the world population has increased every year. Superoxide dismutase 2 (Mn-SOD or SOD2) protects against oxidative stress, a main factor influencing cellular longevity. Polymorphisms in SOD2 have been associated with the development of neurodegenerative diseases, such as Alzheimer's and Parkinson's disease, as well as psychiatric disorders, such as schizophrenia, depression and bipolar disorder. In this study, all of the described natural variants (S10I, A16V, E66V, G76R, I82T and R156W) of SOD2 were subjected to *in silico* analysis using eight different algorithms: SNPeffect, PolyPhen-2, PhD-SNP, PMUT, SIFT, SNAP, SNPs&GO and nsSNPAnalyzer. This analysis revealed disparate results for a few of the algorithms. The results showed that, from at least one algorithm, each amino acid substitution appears to harmfully affect the protein. Structural theoretical models were created for variants through comparative modelling performed using the MHOLline server (which includes MODELLER and PROCHECK) and *ab initio* modelling, using the I-Tasser server. The predicted models were evaluated using TM-align, and the results show that the models were constructed with high accuracy. The RMSD values of the modelled mutants indicated likely pathogenicity for all missense mutations. Structural phylogenetic analysis using ConSurf revealed that human SOD2 is highly conserved. As a result, a human-curated database was generated that enables biologists and clinicians to explore SOD2 nsSNPs, including predictions of their effects and visualisation of the alignment of both the wild-type and mutant structures. The database is freely available at http://bioinfogroup.com/database/ and will be regularly updated.

Introduction

Although aging is a multifactorial process, there is significant evidence that shows that oxidative stress is one of the main factors that influences cellular longevity. Interest in the factors that determine longevity has grown recently because the life expectancy of the world population is increasing. Additionally, in many countries, the main causes of death are currently comorbidities connected to age and oxidative stress.

Superoxide dismutases (SODs) protect against oxidative stress and have three forms: Cu-Zn SOD (SOD1), located in the cytosol; Mn-SOD (SOD2), located in the mitochondrial matrix; and extracellular SOD (SOD3) [1]. The disproportionate rate of intrauterine death and early fatality in Mn-SOD knock-out animals demonstrated the importance of Mn-SOD, rather than SOD1 and SOD3, in foetal development. [1].

The first 24 amino acids of Mn-SOD are the mitochondrial targeting sequence (MTS), which guides and docks the Mn-SOD protein to mitochondria. [1].

Polymorphisms in SOD2 have been associated with the development of neurodegenerative diseases, such as Alzheimer's [2] (A16V) and Parkinson's disease [3] [4] [1] (A16V and I82T), as well as psychiatric disorders, such as schizophrenia [5], depression [6] and bipolar disorder [7]. Similarly, clinical trials showed improvement in symptoms in response to treatment with the glutathione precursor NAC in patients with schizophrenia and

bipolar disorder [8] [9], suggesting that defects in the oxidative stress pathway may contribute to the pathogenesis of various diseases and symptoms. Studies suggest that the effects of all natural variants may primarily reflect functional polymorphism of mitochondrial transport of human MnSOD. As oxidative damage is believed to be an important factor in the pathogenesis of all of these diseases, all of the known variants could possibly contribute to the associated risks. The knowledge of their molecular basis facilitates the diagnosis and design of new drugs.

In this study, we collected the natural variants of SOD2 for *in silico* analysis, which can determine whether these variants influence the protein's three-dimensional structure or stability. Structural theoretical models were created for the variants using comparative modelling performed in MHOLline [10]. MHOLline includes a set of programmes for protein structure analysis, including MODELLER [11] and PROCHECK [12]. I-Tasser [13] was used for *ab initio* modelling. Afterwards, the predicted models were aligned to the wild-type PDB structure using TM-align [14]. Possible effects of the missense variants on protein function could be inferred using bioinformatics tools designed specifically for these types of interpretation, such as PolyPhen-2 [15]. Because of the importance of understanding which variants are disease-related, programmes such as SNPeffect [16], PhD-SNP [17], PMUT [18], SIFT [19], SNAP [20] [21], SNPs&GO [22]

Table 1. Summary of identified SOD2 variants.

Position	Mutation	Feature identifier
10	S10I (S-15I)	VAR_019363
16	A16V (A-9V)	VAR_016183
66	E66V	VAR_019364
76	G76R	VAR_025898
82	I82T	VAR_007165
156	R156W	VAR_019365

doi:10.1371/journal.pone.0065558.t001

and nsSNPAnalyzer [23] were utilised to predict whether a given single-point protein mutation affected the protein function.

As a result, a database was generated for biologists and clinicians to explore SOD2 nsSNPs and the resulting changes in structure and function. This database is freely available at http://bioinfogroup.com/database/and will be regularly updated.

Materials and Methods

Sequence Retrieval

The sequence and natural variants of Mn-SOD were retrieved from the UniProt database.

Non-synonymous SNP Analysis

The functional effects of non-synonymous single-nucleotide substitutions (nsSNPs) were predicted using the following programmes: PhD-SNP [17], PMUT [18], PolyPhen-2 [15], SIFT (Sorting Intolerant from Tolerant) [19], SNAP [20,21], SNPs&GO [22] and nsSNPAnalyzer [23]. SNPeffect [16] was used to evaluate aggregation tendency (TANGO), amyloid propensity (WALTZ), chaperone binding tendency (LIMBO) and protein stability (FoldX).

Comparative and *ab initio* Modelling

The mutant (E66V, G76R, I82T and R156W) models were built using the MHOLline workflow [10] with the crystallographic structure of human SOD2 (PDB ID: 1LUV) as the template. I-Tasser was utilised for the *ab initio* modelling of the S10I and A16V mutants [13]. The TM-scores and root mean square deviations (RMSDs) of the mutant structures with respect to the wild-type structure were calculated using TM-Align [14].

Structural Phylogenetic Analysis

ConSurf was used for high-throughput characterisation of the functional regions in the protein [24]. The degree of conservation of the amino-acid sites among 50 homologues with similar sequences was estimated. The conservation grades were projected onto the molecular surface of the human SOD2 to reveal the patches with highly conserved residues that are often important for biological function.

SOD2 Database Construction

The natural variants listed in the database come from UniProt. For each SNP, we provide predictions of the function effects using SNPeffect, PolyPhen-2, PhD-SNP, PMUT, SIFT, SNAP, SNPs&GO and nsSNPAnalyzer.

The database is web-accessible and can show the following in a comparative table: mutant name; a visualisation of the aligned structures and the predicted functional effects.

Table 2. Predictions of the effect of the missense variations on SOD2 protein function.

Natural Variant	nsSNP Analyzer	PhD-SNP	PMUT	Polyphen-2	SIFT	SNAP	SNPs&GO	TANGO Aggregation Tendency	WALTZ Amyloid Propensity	LIMBO Chaperone Binding Tendency	FoldX Protein Stability
Non synonymous SNP analysis programs											
S10I	Unknown	Neutral	Neutral	Benign	Tolerated	Non-neutral	Disease	Not Affected	Not Affected	Not Affected	Unknown
A16V	Unknown	Neutral	Pathological	Benign	Tolerated	Neutral	Neutral	Not Affected	Not Affected	Not Affected	Unknown
E66V	Disease	Disease	Neutral	Possibly damaging	Tolerated	Neutral	Neutral	Not Affected	Not Affected	Not Affected	Slightly Enhanced
G76R	Disease	Neutral	Pathological	Benign	Tolerated	Non-neutral	Disease	Not Affected	Not Affected	Not Affected	Reduced
I82T	Neutral	Neutral	Neutral	Benign	Affect Protein Function	Non-neutral	Disease	Not Affected	Not Affected	Decreased	Not Affected
R156W	Neutral	Disease	Pathological	Benign	Affect Protein Function	Neutral	Disease	Not Affected	Not Affected	Not Affected	Slyghtly Reduced

doi:10.1371/journal.pone.0065558.t002
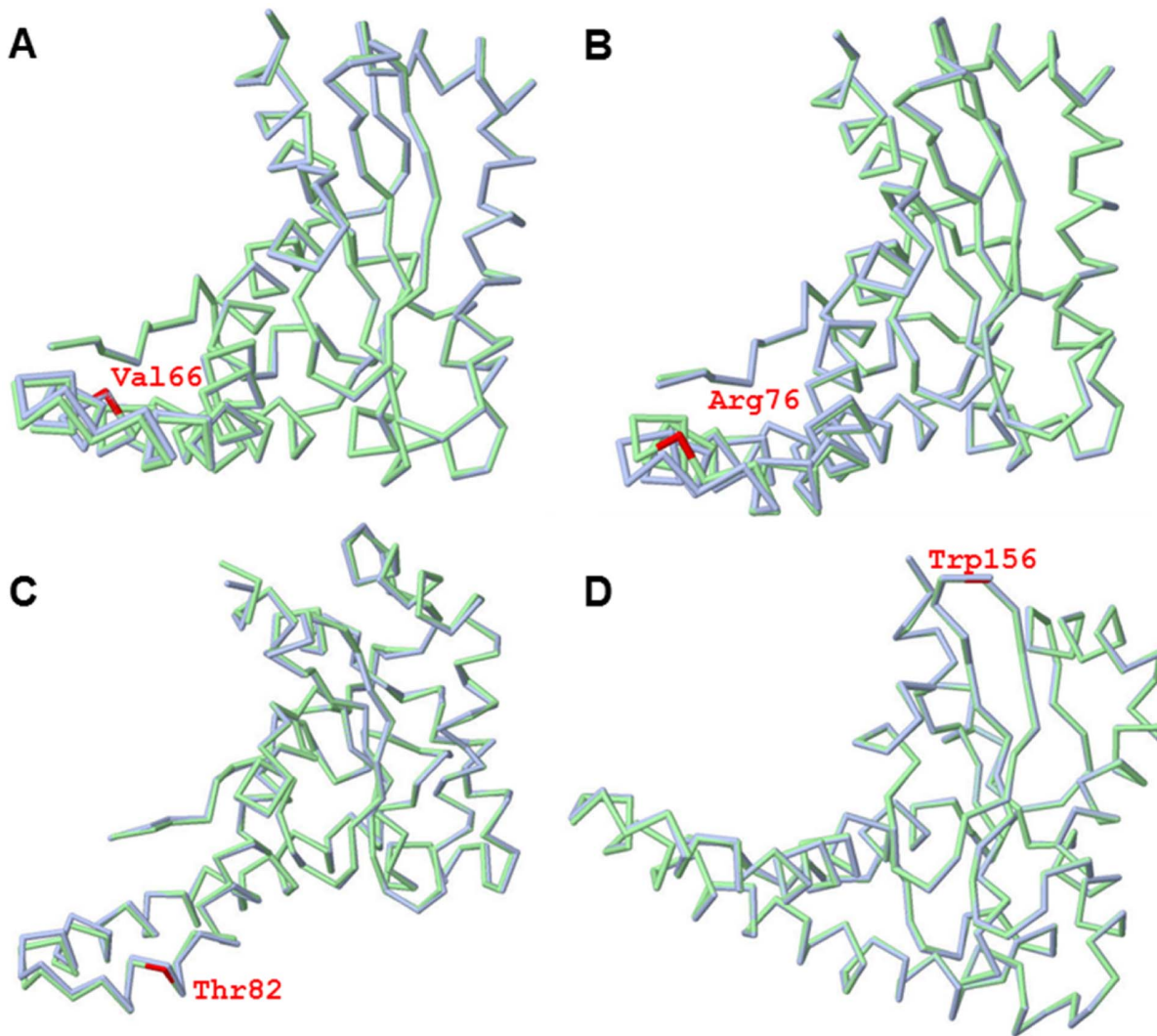
Figure 1. Superimposed native structures (green) and mutant structures (blue) of the SOD2 produced using comparative modelling. A) mutation E66V (E42V), RMSD: 0.21; B) mutation G76R (G52R), RMSD: 0.38; C) mutation I82T (I58T), RMSD: 0.45; D) mutation R156W (R132W), RMSD: 0.16.
doi:10.1371/journal.pone.0065558.g001

Results and Discussion

Sequence Retrieval

The protein sequence and the natural variants of Mn-SOD were retrieved from the UniProt database [25]. The UniProt ID is P04179, and currently, there are natural variants described at six positions. The positions, the substitutions and their references in UniProt are shown in Table 1.

Non-synonymous SNP Analysis

The Mn-SOD variants were subjected to a variety of *in silico* SNP analyses. The results of the non-synonymous SNP analyses are shown in Table 2.

The SNPeffect workflow evaluates aggregation tendency (TANGO), amyloid propensity (WALTZ), chaperone binding tendency (LIMBO) and protein stability (FoldX). The natural variant E66V slightly enhances the protein stability, in contrast with the G76R variant, which reduces the protein stability. The I82T variant decreases the chaperone binding tendency, and the R156W variant slightly reduces the protein stability.

According to PhD-SNP, variants S10I, A16V, G76R and I82T are neutral, whereas variants E66V and R156W cause disease.

The PMUT analysis indicates that the natural variants S10I, E66V and I82T are neutral and that A16V, G76R and R156W are pathological.

The PolyPhen-2 results show that, of the six variants, only E66V may cause damage and that all of the others are benign.

According to SIFT (Sorting Intolerant from Tolerant), tolerance was predicted for the natural variants S10I, A16V, E66V and G76R. I82T and R156W were predicted to affect protein function. The SNAP analysis indicates that variants S10I, G76R and I82T are non-neutral and that A16V, E66V and R156W are neutral.

According to SNPs&GO, variants S10I, G76R, I82T and R156W cause disease, and A16V and E66V are neutral.

The nsSNPAnalyzer results demonstrate that variants S10I and A16V are unknown and variants E66V and G76R cause disease. In contrast, I82T and R156W are neutral.

The SNP analysis, shown in Table 2, indicates that none of the natural variants have only positive results. For each single
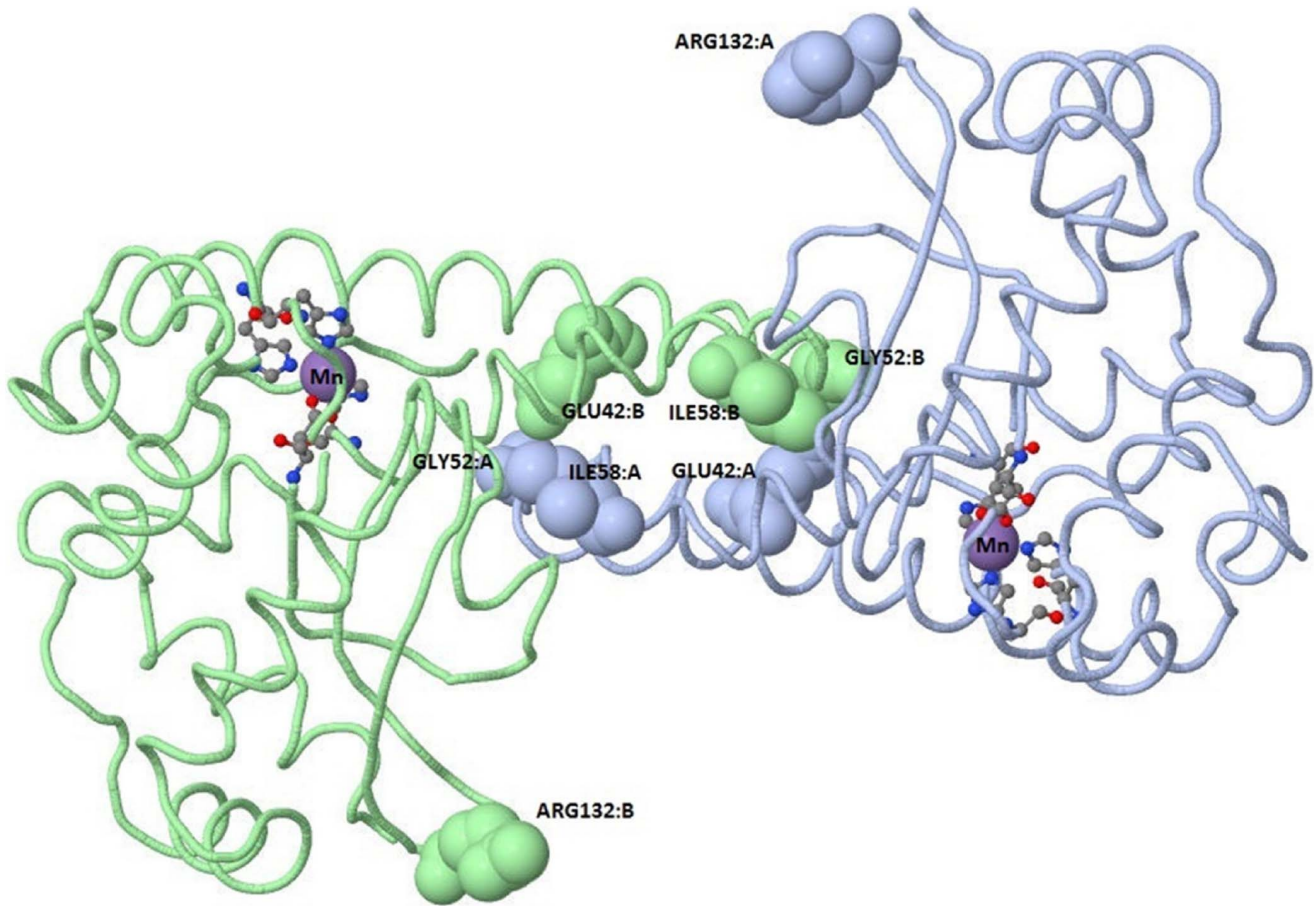
Figure 2. 3D structure of human SOD2 with four missense mutation sites. Two subunits are represented as a backbone in green and blue. Four mutation sites are shown in a sphere representation: E66V, G76R, I82T and R156. The manganese binding site is shown in ball-stick form.
doi:10.1371/journal.pone.0065558.g002

mutation, at least one algorithm indicates a harmful effect on the protein. This result demonstrates the importance of using different algorithms because each algorithm uses different parameters to evaluate the effects of natural variants.

Comparative and *ab initio* Modelling

The natural variants were substituted into the wild-type sequence for comparative modelling. These sequences were submitted to the MHOLline workflow [10]. The theoretical models generated using MHOLline are presented in Figure 1.

Figure 2 shows the two chains of SOD2 (PDB ID: 1LUV), four mutations (the ones that are not in the signal peptide) and the

Table 3. Structure alignment comparing mutant models and wild-type SOD2 models.

Pos.	Variant	TM-Align		
		Align	RMSD	TM-Score
66	E66V (E42V)	1LUV	0.21	0.99834
76	G76R (G52R)	1LUV	0.38	0.995
82	I82T (I58T)	1LUV	0.45	0.995
156	R156W (R132W)	1LUV	0.16	0.995

doi:10.1371/journal.pone.0065558.t003

binding site for manganese. This figure indicates that 3 of the variants localise in the interaction surfaces of chains A and B. This localisation may adversely influence dimer formation, especially the I58T mutation, which affects the stability of the tetrameric (dimer-dimer) interface [26].
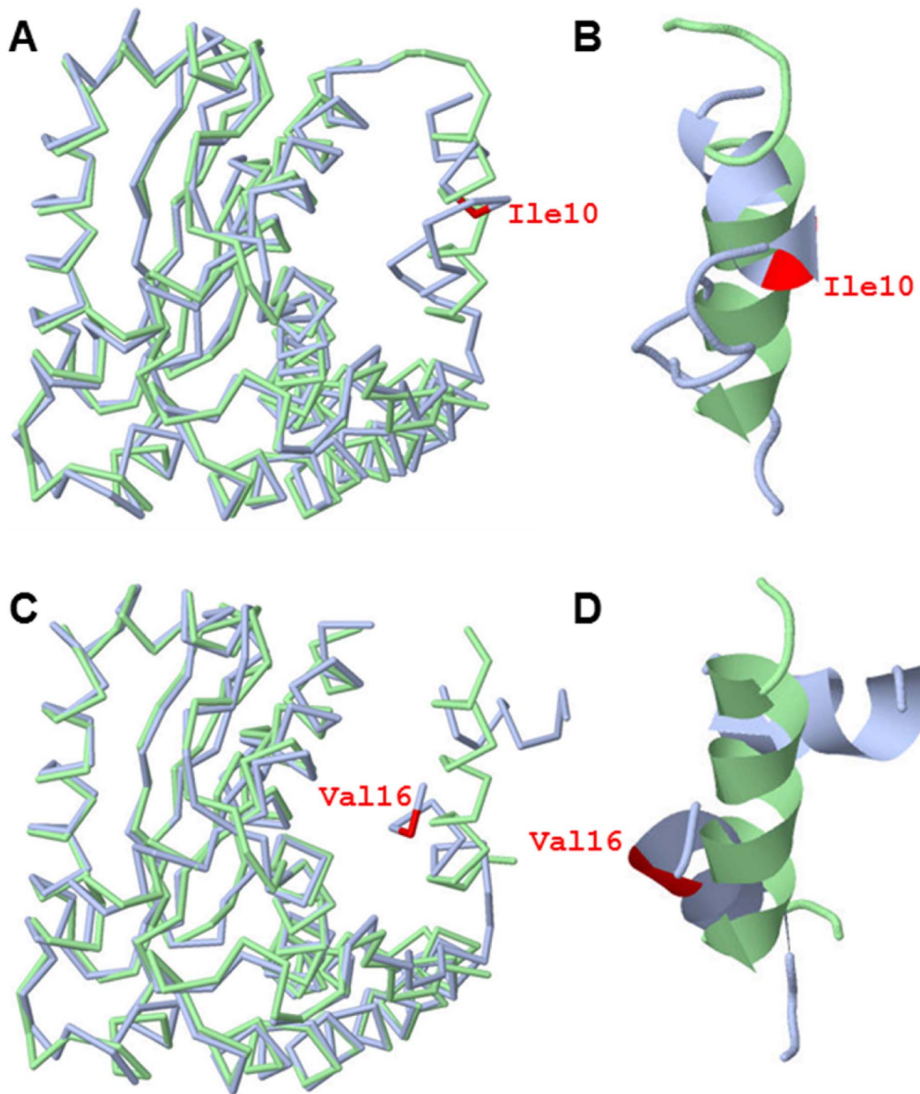
An alignment between the native and mutant structures was performed using TM-Align [14]. Parameters such as the TM-score and root mean square deviation (RMSD) were used to analyse the topology and structural similarity of the models. TM-score was used to assess the topological similarity of two protein structures, while RMSD was the measure of the average distance between the backbones of the superimposed proteins [27]. The RMSD values for the modelled mutants were significant for pathogenicity for all missense mutations (Figure 1 and Table 3). RMSD values greater than 0.15 were considered significant structural perturbations that could have functional implications for the protein [28].

To analyse the three-dimensional effects of the S10I and A16V mutations, which are located in the signal peptide, *ab initio* modelling was necessary because the signalling sequence cannot be resolved experimentally. The I-Tasser server [13] was utilised for the *ab initio* modelling. As shown in Figure 3 and Table 4, the structural alignment of the *ab initio* mutant models and the *ab initio* native models reveals that the S10I and A16V mutations exhibited high RMSD values and disrupted the alpha helix in the signal peptide.

Figure 3. Superimposed native structures (green) and mutant structures (blue) of the SOD2 produced using *ab initio* modelling. A) S10I (S-15I) mutation highlighted in red. B) This mutation disrupts the alpha helix, RMSD: 2.02. C) A16V (A-9V) mutation highlighted in red. D) This mutation disrupts the alpha helix, RMSD: 1.94.
doi:10.1371/journal.pone.0065558.g003

Structural Phylogenetic Analysis

The ConSurf [24] results are based on the concept of identify functional regions in proteins, taking into account by considering the evolutionary relationships among their sequence homologues. An advantage of ConSurf over other methods is the accurate computation of the evolutionary rate using either an empirical

Table 4. Structure alignment of *ab initio* SOD2 mutant models with the *ab initio* wild-type model.

Pos.	Variant	I-Tasser			TM-Align	
		C-score	TM-score	RMSD	RMSD	TM-Score
10	S10I (S-15I)	0.18	0.69±0.12	5.9±3.7	2.02	0.90520
16	A16V (A-9V)	0.15	0.69±0.12	5.9±3.7	1.94	0.91721

doi:10.1371/journal.pone.0065558.t004

Bayesian method or a maximum likelihood method. Thus, ConSurf can correctly discriminate between the conservation caused by a short evolutionary time and genuine sequence conservation. The surface residues with the most variation are depicted in blue, and the conserved residues are depicted in purple in the protein structures (Figure 4). Our findings revealed that human SOD2 is highly conserved (Figure 4). The sequence alignment of the SOD2 from various species (Figure 5) reveals that residues E66 and G76 are conserved, whereas I82 and R156 are variable.
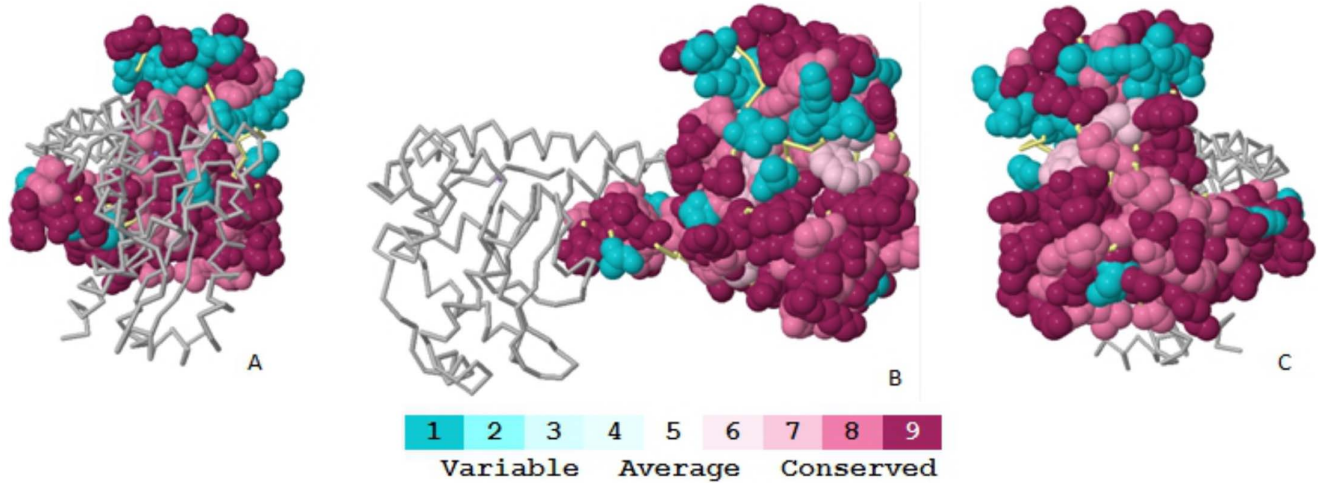
The conservation analysis of ConSurf used the evolutionary conservation scores of the residues to identify functional regions from proteins with known three-dimensional structures. The degree of conservation of the amino acid sites among the nine homologues with similar sequences (Figure 5) was estimated. The conservation grades were projected onto the molecular surface of the proteins to reveal the patches of highly conserved residues that are often important for biological function. Mutations E66 and G76 are conserved, whereas mutations I82 and R156 are variable.

Figure 4. Conservation profile of the Mn-SOD (PDB ID: 1LUV) using ConSurf conservational analysis. Mn-SOD is represented as a spacefill model, where the residue conservation scored is colour-coded onto the surface. The backbone model represents the other chain of a Mn-SOD dimer, chain B. The colour-coding bar shows the colouring scheme: conserved amino acids are coloured bordeaux, residues with average conservation are white, and variable amino acids are turquoise.
doi:10.1371/journal.pone.0065558.g004
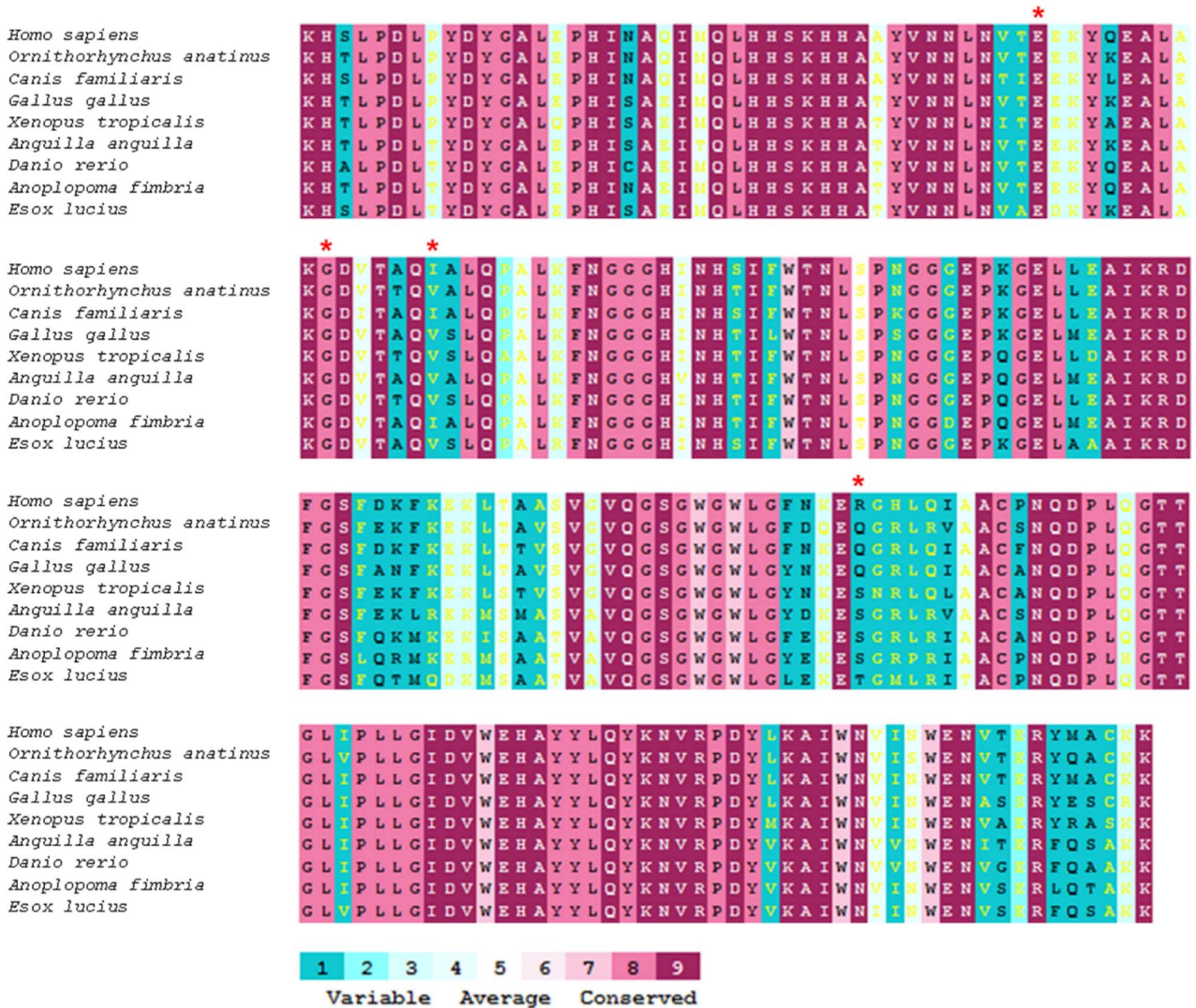
Figure 5. Multiple protein sequence alignment using ConSurf shows evolutionary conservation of amino acid residues. The colour-coding bar shows the colouring scheme: conserved amino acids are coloured bordeaux, residues of average conservation are white, and variable amino acids are turquoise. SNP positions are marked by an asterisk.
doi:10.1371/journal.pone.0065558.g005

6
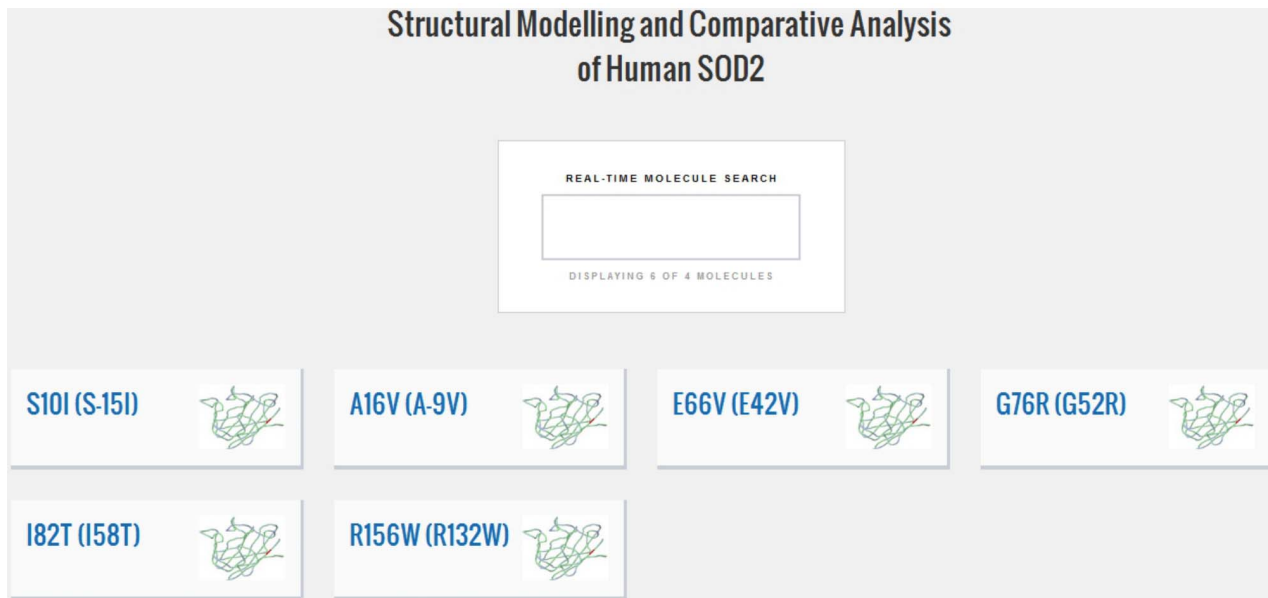
Figure 6. Screenshot of the SOD2 Database web interface for structural modelling and comparative analysis.
doi:10.1371/journal.pone.0065558.g006

Generally, residues that are implicated in biological processes, such as those located in active sites, involved in protein-protein or protein-ligand interactions, or implicated in protein structure and folding stability, are subject to greater selective pressure and are usually more conserved than other residues.

SOD2 Database

The SOD2 database currently contains all of the natural variants listed in UniProt. For each SNP, we provide the predictions of functional effects, indicated as Disease/Pathological or Neutral/Tolerated, from SNPeffect, PolyPhen-2, PhD-SNP, PMUT, SIFT, SNAP, SNPs&GO and nsSNPAnalyzer.

The database interface (Figure 6) allows users to search for a mutation by its non-synonymous SNP.

The database is curated by humans and will be updated as new natural variants are discovered.

The SOD2 database allows a user to quickly retrieve and rapidly analyse the predicted effects of protein variants. In addition to predicting the effects of variants, an alignment of the wild-type and mutant structures can be visualised using the database.

The major feature that distinguishes the SOD2 database from other databases is that this database can use predictions from several algorithms for all of the known natural variants of Mn-SOD. Furthermore, the user has access to an alignment of the wild type and mutant structures and can thus visualise the damage that a SNP can cause. Our ultimate goal is to turn the database into a toolbox for researchers studying this protein. The *in silico* analysis of Mn-SOD in this database will help in the design and prioritisation of further experimental research.

Author Contributions

Conceived and designed the experiments: MDCC JFM. Performed the experiments: MDCC. Analyzed the data: MDCC JFM. Contributed reagents/materials/analysis tools: JFM. Wrote the paper: MDCC JFM.

References

1. Wang V, Chen SY, Chuang TC, Shan DE, Soong BW, et al. (2010) Val-9Ala and Ile+58Thr polymorphism of MnSOD in Parkinson's disease. Clin Biochem 43: 979–982.
2. Wiener HW, Perry RT, Chen Z, Harrell LE, Go RC (2007) A polymorphism in SOD2 is associated with development of Alzheimer's disease. Genes Brain Behav 6: 770–775.
3. Shimoda-Matsubayashi S, Matsumine H, Kobayashi T, Nakagawa-Hattori Y, Shimizu Y, et al. (1996) Structural dimorphism in the mitochondrial targeting sequence in the human manganese superoxide dismutase gene. A predictive evidence for conformational change to influence mitochondrial transport and a study of allelic association in Parkinson's disease. Biochem Biophys Res Commun 226: 561–565.
4. Singh M, Khan AJ, Shah PP, Shukla R, Khanna VK, et al. (2008) Polymorphism in environment responsive genes and association with Parkinson disease. Mol Cell Biochem 312: 131–138.
5. Akyol O, Yanik M, Elyas H, Namli M, Canatan H, et al. (2005) Association between Ala-9Val polymorphism of Mn-SOD gene and schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry 29: 123–131.
6. Galecki P, Smigielski J, Florkowski A, Bobinska K, Pietras T, et al. (2010) Analysis of two polymorphisms of the manganese superoxide dismutase gene (Ile-58Thr and Ala-9Val) in patients with recurrent depressive disorder. Psychiatry Res 179: 43–46.
7. Fullerton JM, Tiwari Y, Agahi G, Heath A, Berk M, et al. (2010) Assessing oxidative pathway genes as risk factors for bipolar disorder. Bipolar Disord 12: 550–556.
8. Berk M, Copolov D, Dean O, Lu K, Jeavons S, et al. (2008) N-acetyl cysteine as a glutathione precursor for schizophrenia–a double-blind, randomized, placebo-controlled trial. Biol Psychiatry 64: 361–368.
9. Berk M, Copolov DL, Dean O, Lu K, Jeavons S, et al. (2008) N-acetyl cysteine for depressive symptoms in bipolar disorder–a double-blind randomized placebo-controlled trial. Biol Psychiatry 64: 468–475.
10. Capriles PV, Guimaraes AC, Otto TD, Miranda AB, Dardenne LE, et al. (2010) Structural modelling and comparative analysis of homologous, analogous and specific proteins from Trypanosoma cruzi versus Homo sapiens: putative drug targets for chagas' disease treatment. BMC Genomics 11: 610.
11. Sanchez R, Sali A (1997) Evaluation of comparative protein structure modeling by MODELLER-3. Proteins Suppl 1: 50–58.
12. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26: 283–291.
13. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725–738.
14. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33: 2302–2309.

15. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249.
16. De Baets G, Van Durme J, Reumers J, Maurer-Stroh S, Vanhee P, et al. (2012) SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. Nucleic Acids Res 40: D935–939.
17. Capriotti E, Calabrese R, Casadio R (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 22: 2729–2734.
18. Ferrer-Costa C, Orozco M, de la Cruz X (2002) Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. J Mol Biol 315: 771–786.
19. Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11: 863–874.
20. Bromberg Y, Rost B (2007) SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res 35: 3823–3835.
21. Bromberg Y, Yachdav G, Rost B (2008) SNAP predicts effect of mutations on protein function. Bioinformatics 24: 2397–2398.
22. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 30: 1237–1244.
23. Bao L, Zhou M, Cui Y (2005) nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. Nucleic Acids Res 33: W480–482.
24. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 38: W529–533.
25. UNIPROT (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 40: D71–75.
26. Borgstahl GE, Parge HE, Hickey MJ, Johnson MJ, Boissinot M, et al. (1996) Human mitochondrial manganese superoxide dismutase polymorphic variant Ile58Thr reduces activity by destabilizing the tetrameric interface. Biochemistry 35: 4287–4297.
27. Jimenez-Lopez JC, Gachomo EW, Seufferheld MJ, Kotchoni SO (2010) The maize ALDH protein superfamily: linking structural features to functional specificities. BMC Struct Biol 10: 43.
28. Mistri M, Tamhankar PM, Sheth F, Sanghavi D, Kondurkar P, et al. (2012) Identification of novel mutations in HEXA gene in children affected with Tay Sachs disease from India. PLoS One 7: e39122.